

Safe and Economic Re-Use of Ontologies: A Logic-Based Methodology and Tool Support

Ernesto Jiménez-Ruiz¹ Bernardo Cuenca Grau²
Ulrike Sattler³ *Thomas Schneider*³ Rafael Berlanga¹

¹Computer Languages and Systems, Universitat Jaume I, Spain

²Computing Laboratory, University of Oxford, UK

³Computer Science, University of Manchester, UK

DL, 13–16 May 2008

Our approach in a nutshell

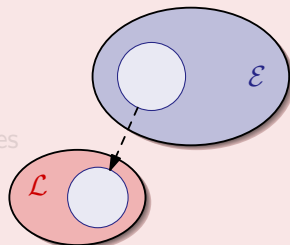
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

- 1 Don't change their meaning! ✓

Economic import of the external ontologies

- 2 Import only the relevant parts ...
- 3 ... without loss of information! ✓



- Tool support — Protégé plugin
- Work in progress!

Our approach in a nutshell

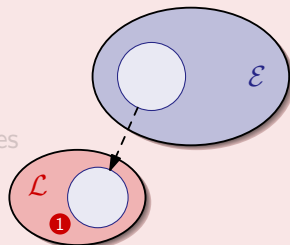
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

- 1 Don't change their meaning! ✓

Economic import of the external ontologies

- 2 Import only the relevant parts ...
- 3 ... without loss of information! ✓



- Tool support — Protégé plugin
- Work in progress!

Our approach in a nutshell

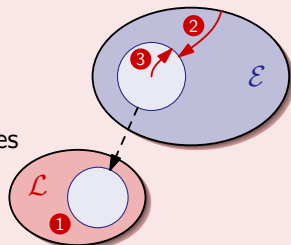
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

- 1 Don't change their meaning! ✓

Economic import of the external ontologies

- 2 Import only the relevant parts ...
- 3 ... without loss of information! ✓



- Tool support — Protégé plugin
- Work in progress!

Our approach in a nutshell

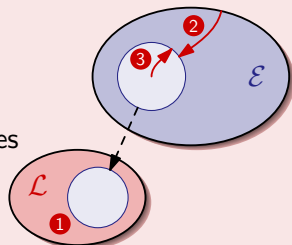
Logic-based methodology for the re-use of ontologies

Safe use of imported symbols

- 1 Don't change their meaning! ✓

Economic import of the external ontologies

- 2 Import only the relevant parts ...
- 3 ... without loss of information! ✓

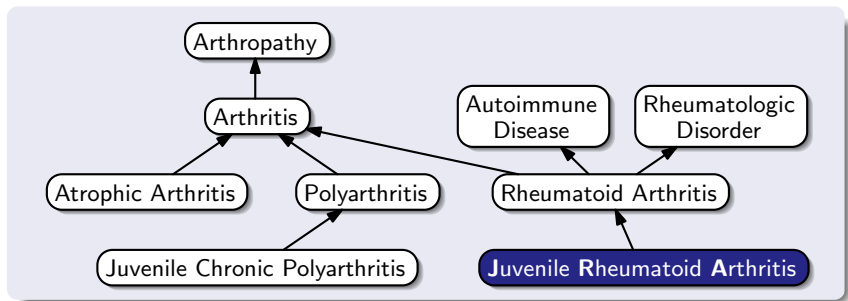


- Tool support — Protégé plugin
- Work in progress!

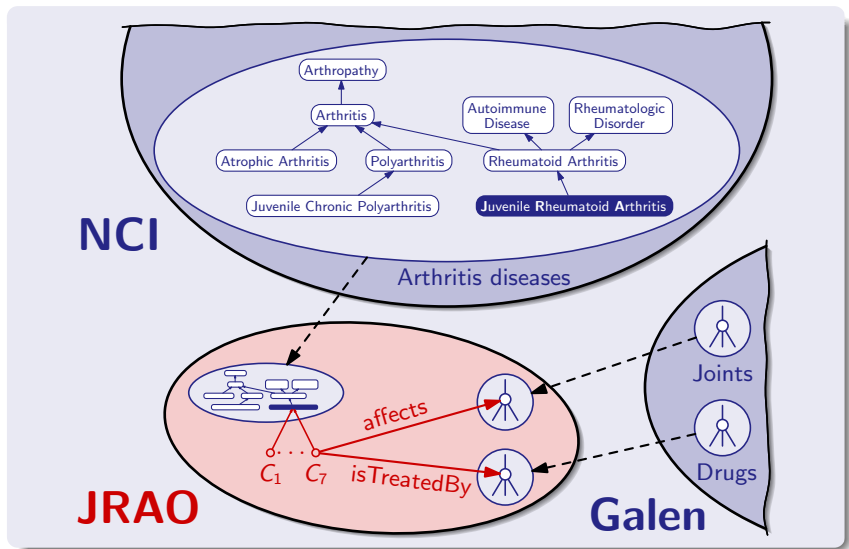
And now . . .

- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support
- 4 Perspectives

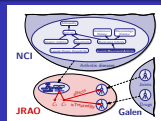
A re-use scenario: the *Health-e-Child* project



A re-use scenario: the *Health-e-Child* project



A case for safe and economic re-use



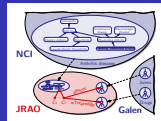
Reasons for re-use

- Saves time for re-writing
- Provides access to well-established knowledge
- Doesn't require expertise in drugs, proteins, anatomy etc.

Guarantees to provide

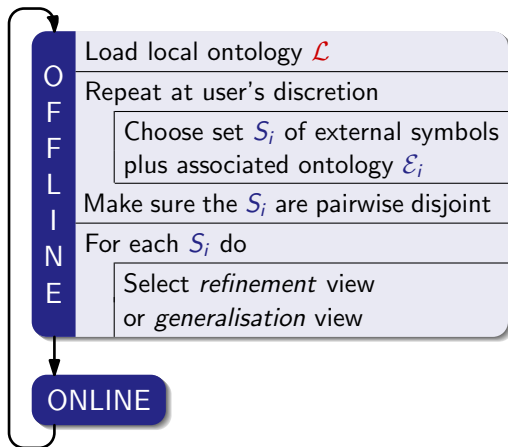
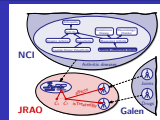
- **[safe]** Importing terms doesn't change their meaning.
- **[eco]** Import all relevant parts of external ontologies.
- **[aux]** The order of imports doesn't matter.

And now . . .

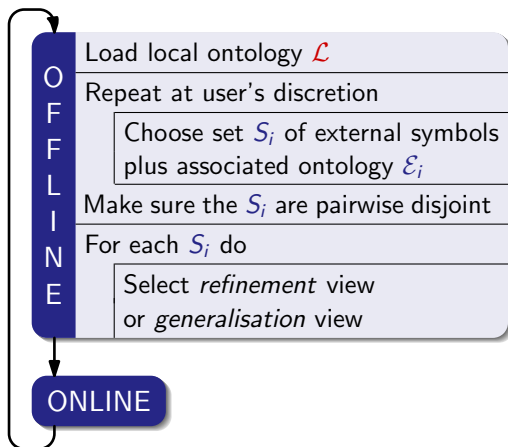
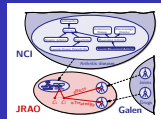


- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support
- 4 Perspectives

A working cycle: the offline phase

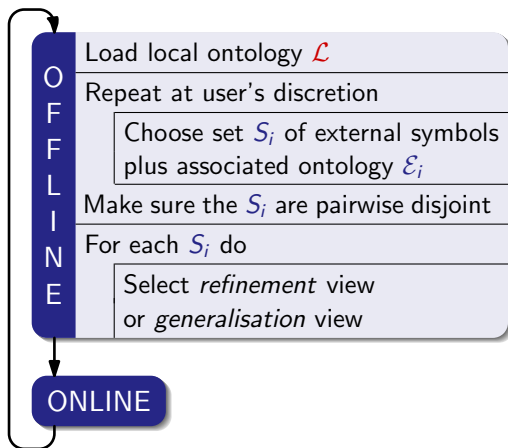
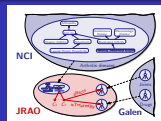


A working cycle: the offline phase



$$S_1 = \{ \text{JRAO} \} \quad \mathcal{E}_1 = \text{NCI}$$

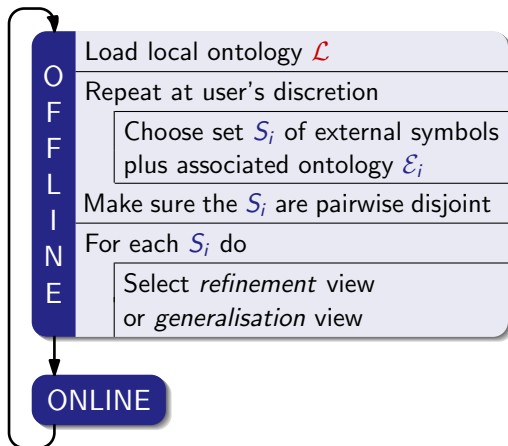
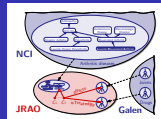
A working cycle: the offline phase



$$S_1 = \{JRA\} \quad \mathcal{E}_1 = \mathbf{NCI}$$

$$S_2 = \{KneeJoint, Fever\} \\ \mathcal{E}_2 = \mathbf{Galen}$$

A working cycle: the offline phase



$$S_1 = \{JRA\} \quad \mathcal{E}_1 = \mathbf{NCI}$$

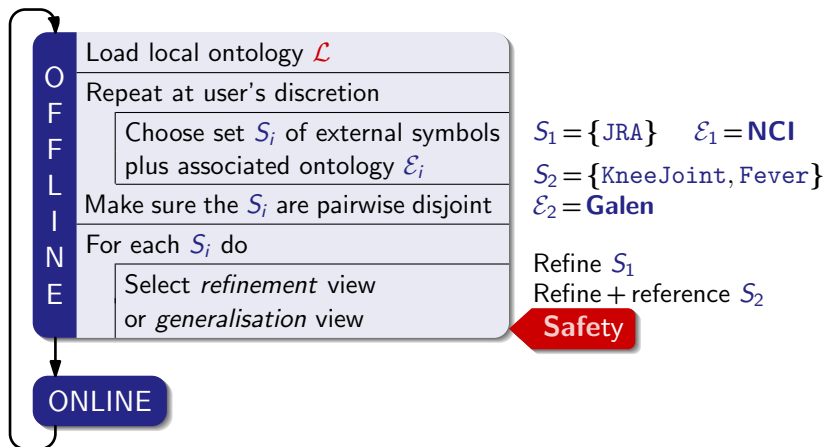
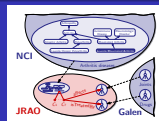
$$S_2 = \{KneeJoint, Fever\}$$

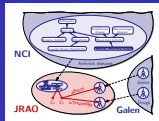
$$\mathcal{E}_2 = \mathbf{Galen}$$

Refine S_1

Refine + reference S_2

A working cycle: the offline phase





Formalising the *Safety Guarantee*

Safety

Importing terms doesn't change their meaning.

Example

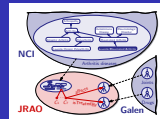
$\mathbf{JRAO} \cup \mathbf{NCI} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}$
 iff $\mathbf{NCI} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}.$

Definition (Safety)

\mathcal{L} guarantees safety if for every $i = 1, \dots, n$:

For every \mathcal{E}'_i with $\text{Sig}(\mathcal{L}) \cap \text{Sig}(\mathcal{E}'_i) \subseteq S_i$,

$\mathcal{L} \cup \mathcal{E}'_i$ is a conservative extension of \mathcal{E}'_i .



Formalising the *Safety Guarantee*

Safety

Importing terms doesn't change their meaning.

Example

$\mathbf{JRAO} \cup \mathbf{NCI} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}$
 iff $\mathbf{NCI} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}.$

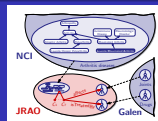
Definition (Safety)

\mathcal{L} guarantees safety if for every $i = 1, \dots, n$:

For every \mathcal{E}'_i with $\mathbf{Sig}(\mathcal{L}) \cap \mathbf{Sig}(\mathcal{E}'_i) \subseteq S_i$,

$\mathcal{L} \cup \mathcal{E}'_i$ is a conservative extension of \mathcal{E}'_i .

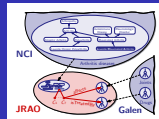
Approximating conservativity



$\mathcal{L} \cup \mathcal{E}'_i$ is a *deductive conservative extension* of \mathcal{E}'_i

$\forall \alpha$ with $\text{Sig}(\alpha) \subseteq \text{Sig}(\mathcal{E}'_i)$, $\mathcal{L} \cup \mathcal{E}'_i \models \alpha$ iff $\mathcal{E}'_i \models \alpha$.

Approximating conservativity



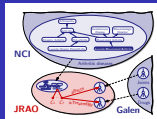
$\mathcal{L} \cup \mathcal{E}'_i$ is a *deductive conservative extension* of \mathcal{E}'_i

$\forall \alpha$ with $\text{Sig}(\alpha) \subseteq \text{Sig}(\mathcal{E}'_i)$, $\mathcal{L} \cup \mathcal{E}'_i \models \alpha$ iff $\mathcal{E}'_i \models \alpha$.

↑

$\mathcal{L} \cup \mathcal{E}'_i$ is a *model-theoretic conservative extension* of \mathcal{E}'_i

$\forall \mathcal{I}$ for \mathcal{E}'_i $\exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i$: $\mathcal{I}|_{\text{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\text{Sig}(\mathcal{E}'_i)}$.



Approximating conservativity

$\mathcal{L} \cup \mathcal{E}'_i$ is a *deductive conservative extension* of \mathcal{E}'_i

$\forall \alpha$ with $\text{Sig}(\alpha) \subseteq \text{Sig}(\mathcal{E}'_i)$, $\mathcal{L} \cup \mathcal{E}'_i \models \alpha$ iff $\mathcal{E}'_i \models \alpha$.

↑

$\mathcal{L} \cup \mathcal{E}'_i$ is a *model-theoretic conservative extension* of \mathcal{E}'_i

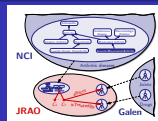
$\forall \mathcal{I}$ for \mathcal{E}'_i $\exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i$: $\mathcal{I}|_{\text{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\text{Sig}(\mathcal{E}'_i)}$.

↑

\mathcal{L} is *semantically \perp -local* w.r.t. $\mathcal{L} \cup \mathcal{E}'_i$

$\forall \mathcal{I}$ for \mathcal{E}'_i $\exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i$: $\mathcal{I}|_{\text{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\text{Sig}(\mathcal{E}'_i)}$
such that $X^{\mathcal{J}} = \emptyset$ for each $X \in \text{Sig}(\mathcal{L}) \setminus S_i$.

Approximating conservativity



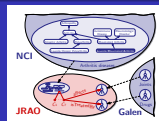
\mathcal{L} is *semantically \perp -local* w.r.t. $\mathcal{L} \cup \mathcal{E}'_i$

$\forall \mathcal{I}$ for $\mathcal{E}'_i \quad \exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i \quad : \quad \mathcal{I}|_{\text{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\text{Sig}(\mathcal{E}'_i)}$

such that $X^{\mathcal{J}} = \emptyset$ for each $X \in \text{Sig}(\mathcal{L}) \setminus S_i$.

*

Approximating conservativity



\mathcal{L} is *semantically* \perp -local w.r.t. $\mathcal{L} \cup \mathcal{E}'_i$

$\forall \mathcal{I}$ for \mathcal{E}'_i $\exists \mathcal{J}$ for $\mathcal{L} \cup \mathcal{E}'_i$: $\mathcal{I}|_{\text{Sig}(\mathcal{E}'_i)} = \mathcal{J}|_{\text{Sig}(\mathcal{E}'_i)}$

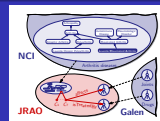
such that $X^{\mathcal{J}} = \emptyset$ for each $X \in \text{Sig}(\mathcal{L}) \setminus S_i$. *

↑

\mathcal{L} is *syntactically* \perp -local w.r.t. S_i

- all GCIs in \mathcal{L} are of the form $C_{\perp} \sqsubseteq C$ or $C \sqsubseteq C_{\top}$ where $C_{\perp}^{\mathcal{I}} = \emptyset$ and $C_{\top}^{\mathcal{I}} = \Delta^{\mathcal{I}}$ follow from *
- similar conditions for RIs and $\text{Trans}(R)$ statements

Providing safety



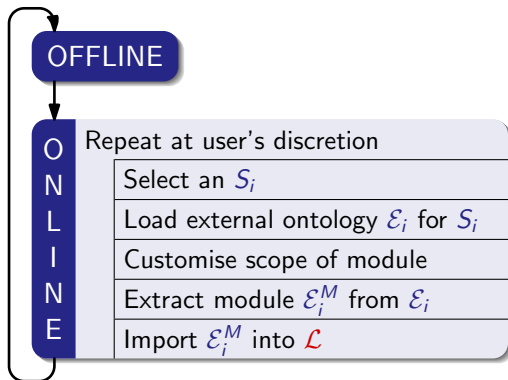
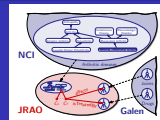
Examples

- $C_7 \sqsubseteq JRA$ ✓ \perp -local
- $GeneticDisorder \sqsubseteq C_7$ ✓ T-local
- $JRA \sqsubseteq GeneticDisorder$ ✗ non-local

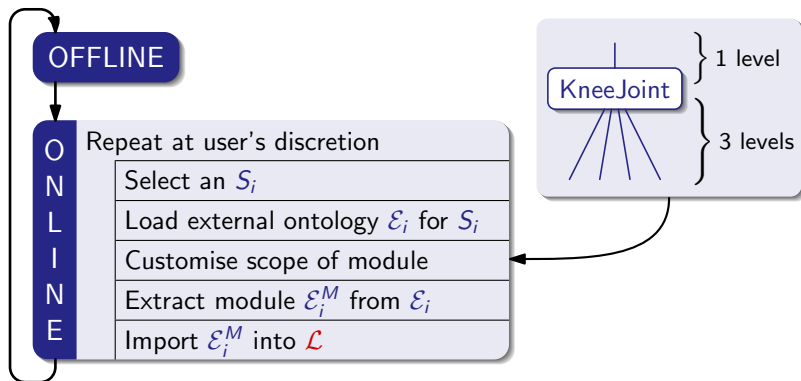
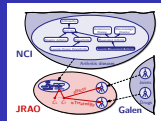
Theorem [Cuenca Grau, Horrocks, Kazakov, Sattler 2007]

If \mathcal{L} is syntactically local w.r.t. each S_i , then \mathcal{L} guarantees safety.

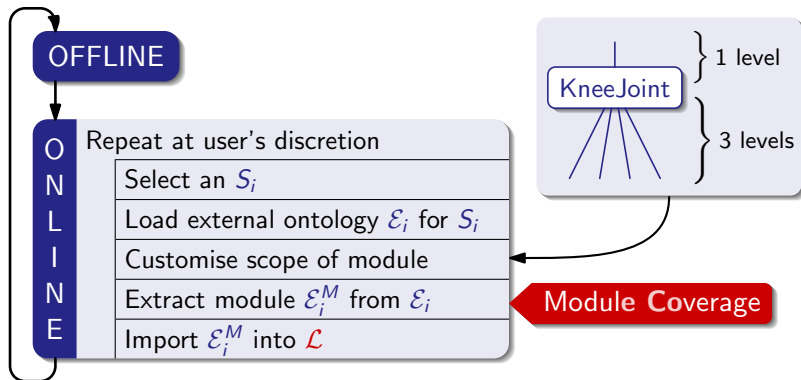
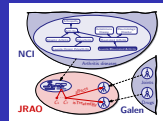
The online phase



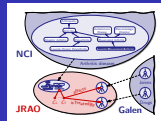
The online phase



The online phase



Formalising the *Module Coverage Guarantee*



Module coverage

Import all relevant parts of external ontologies.

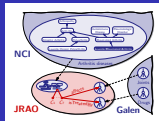
Example

$\mathbf{JRAO} \cup \mathbf{NCI} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}$
 iff $\mathbf{JRAO} \cup \mathbf{NCI\text{-}module} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}.$

Definition (Module coverage)

Let $\mathcal{E}_i^M \subseteq \mathcal{E}_i$ with $S_i \subseteq \text{Sig}(\mathcal{E}_i^M)$. \mathcal{E}_i^M guarantees coverage of S_i if:

For every \mathcal{L}' with $\text{Sig}(\mathcal{L}') \cap \text{Sig}(\mathcal{E}_i) \subseteq S_i$,
 $\mathcal{L}' \cup \mathcal{E}_i$ is a conservative extension of $\mathcal{L}' \cup \mathcal{E}_i^M$.



Formalising the *Module Coverage Guarantee*

Module coverage

Import all relevant parts of external ontologies.

Example

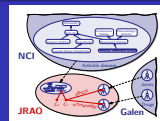
$\mathbf{JRAO} \cup \mathbf{NCI} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}$
 iff $\mathbf{JRAO} \cup \mathbf{NCI\text{-}module} \models \mathbf{JRA} \sqsubseteq \mathbf{GeneticDisorder}.$

Definition (Module coverage)

Let $\mathcal{E}_i^M \subseteq \mathcal{E}_i$ with $S_i \subseteq \text{Sig}(\mathcal{E}_i^M)$. \mathcal{E}_i^M guarantees coverage of S_i if:

For every \mathcal{L}' with $\text{Sig}(\mathcal{L}') \cap \text{Sig}(\mathcal{E}_i) \subseteq S_i$,
 $\mathcal{L}' \cup \mathcal{E}_i$ is a conservative extension of $\mathcal{L}' \cup \mathcal{E}_i^M$.

Providing coverage



- Coverage is again provided using locality.
- Locality-based modules = syntactic approximations of conservativity-based modules
 - in general not minimal
 - efficiently computable

And now . . .

- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support**
- 4 Perspectives

A Protégé Plugin for the reuse of Ontologies: Safe and Économique

http://semanticweb.org/Ontology1206804237717.owl - [/Users/schneidt/Documents/DLista/Tool-Modularity/Experiments/HeC_Use_Cases/JRAO.owl]

Ontology1206804237717.owl http://semanticweb.org/Ontology1206804237717.owl

Active Ontology Entites Classes Object Properties Data Properties Individuals OWLViz DL Query ProSE Manager

ProSE External Ontology URIs

External URI Annotations

label
 ExternalURI: http://www.mindswap.org/2003/CancerOntology/nciOncology.owl Bottom Locality: true

label
 ExternalURI: http://krono.act.uji.es/Links/ontologies/galen.owl Bottom Locality: true

ProSE Safe Protege Manager:

Select Signature Group

External Signature Groups

- http://www.mindswap.org/2003/CancerOntology/nciOncology.owl
- http://krono.act.uji.es/Links/ontologies/galen.owl
- No defined External URI

Locality Type:

Bottom Locality ▲ Non-Local Axioms Detected

Use of External Signature

Refine External Signature

Only Reference External Signature

Generalize External Signature

Importing Information

Selected External Signature will be used for refinement. An Upper Module is extracted.

Extension of the selected Signature

Signature from: /www.mindswap.org/2003/CancerOntology/nciOncology.owl

Extend Signature with Subclasses. Levels: 1

Extend Signature with Superclasses. Levels: 1

★ The signature has been extended

ProSE External-Local Class Hierarchy

- Thing
 - Anti-DNA_Antibody
 - Azathioprine
 - Cisplatin_Cyclosporine
 - Cyclosporine
 - Erosion
 - Etanercept
 - Hemoglobin
 - Infliximab
 - Interleukin_Gene
 - LymphocyteCount
 - Methotrexate
 - NeutrophilCount
 - Nonsteroidal_AntiInflammatory_Drug
 - Oedema
 - Juvenile_Rheumatoid_Arthritis
 - Pain
 - PlateletCount
 - Prednisone
 - Rheumatoid_Factor
 - SynoviaFluid
 - SynoviaJoint
 - Tumor_Necrosis_Factor_Family_Gene
 - WestergrenESRProcedure

Non-Local Axioms for selected External Signature

Non-Local Axioms

- Juvenile_Rheumatoid_Arthritis subClassOf Oedema

Importing-Preview Actions

URI for Module: no.act.uji.es/Links/ontologies/module1206908087974.owl

Importing-Preview Actions

- Extract Module
 - Preview Module
 - Import Module
- Import Whole Ontology

Module Information

Number of Axioms (Module/Ontology): 398 / 395124

Number of Classes (Module/Ontology): 226 / 27652

Number of Properties (Module/Ontology): 13 / 70

Number of Individuals (Module/Ontology): 0 / 0

And now . . .

- 1 Why ontology re-use?
- 2 A safe and economic methodology
- 3 Tool support
- 4 Perspectives**

“Shopping for symbols”

Extend module scope customisation:



- Browse external ontology and pick symbols
- At each stage, view resulting module
- “Check out” module

→ Treemaps?

Other plans

- Optimise module extraction
- Import “by reference” as opposed to “by value”
- Multi-user scenario
- Module extraction service at owl.cs.manchester.ac.uk
- Modularity tool tutorial at ISWC 2008
- Perform user study and improve interface

Other plans

- Optimise module extraction
- Import “by reference” as opposed to “by value”
- Multi-user scenario
- Module extraction service at owl.cs.manchester.ac.uk
- Modularity tool tutorial at ISWC 2008
- Perform user study and improve interface

Other plans

- Optimise module extraction
- Import “by reference” as opposed to “by value”
- Multi-user scenario
- Module extraction service at owl.cs.manchester.ac.uk
- Modularity tool tutorial at ISWC 2008
- Perform user study and improve interface

Invitation

We want you...

- ... to work with us on incorporating our services into your workflows!
- ... r favourite ontologies and real-life signatures!



Contact

`schneider@cs.man.ac.uk`

Invitation

We want you...

- ... to work with us on incorporating our services into your workflows!
- ... our favourite ontologies and real-life signatures!



Contact

schneider@cs.man.ac.uk

Thank you!

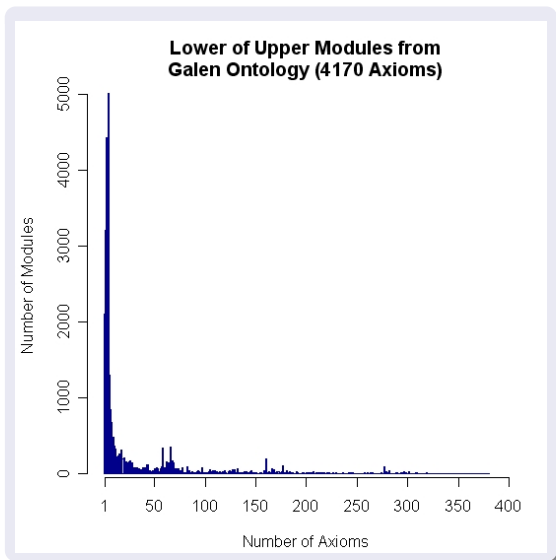
“Synthetic” experiments

Setting

- Randomly generated signatures of size 1 . . . 330
- Computed *Lower of Upper Module (LUM)* for each such signature

Results

- 99 % of **Galen** LUMs contain < 5 % of **Galen**'s axioms
- similar findings for **NCI**



“Real-life” experiments

Setting

LUMs for manually selected signatures from **Galen** and **NCI**
(*Health-e-Child* context: JRA + Cardiomyopathies)

Results

Ext. Ont.	# Sig.	# axioms	
Galen	11	105	(2.5 %)
Galen	72	620	(14.9 %)
Galen	76	736	(17.6 %)
NCI	18	488	(0.1 %)
NCI	124	4751	(1.2 %)
NCI	144	5057	(1.3 %)

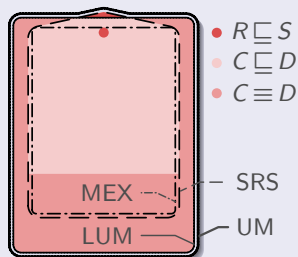
Comparing experiments

Setting

- SNOMED (health care; restricted language; 350,000 axioms)
- Initial signatures: terms from intensive care unit
- Compared UM, LUM to MEX (conservativity-based modules) and SRS (Seidenberg/Rector segments)

Results

# Sig.	# axioms in %		
	MEX	SRS	(L)UM
4,000	2	2	4
16,000	7	7	10
24,000	10	10	15
time	4–5 s	1 s	4–7 s



More links

Protégé and ProSÉ

- protege.stanford.edu
- krono.act.uji.es/people/Ernesto/safety-ontology-reuse

Health-e-Child

- www.health-e-child.org

NCI and Galen

- nciterms.nci.nih.gov/NCIBrowser/Dictionary.do
- ftp1.nci.nih.gov/pub/cacore/EVS/NCIThesaurus
- www.co-ode.org/galen